

DOI: 10.1111/hequ.12185

WILEY Higher Education Quarterly

## ARTICLE - SPECIAL ISSUE

# Do performance-based research funding systems affect research production and impact?

Daniele Checchi<sup>1,2</sup>  | Marco Malgarini<sup>2</sup> | Scipione Sarlo<sup>2</sup><sup>1</sup>Department of Economics, University of Milan, Milan, Italy<sup>2</sup>ANVUR (Agenzia Nazionale per la Valutazione dell'Università e della Ricerca), Rome, Italy

## Abstract

We study the potential impact of introducing performance-based funding systems (PBFS) on national research systems, using information on the number of publications and their scientific impact (citations or publications in top-ranked journals) for 31 countries over the period 1996–2016. The analysis is performed both at the aggregate level and looking separately at each of the six main scientific areas identified according to the Organisation for Economic Co-operation and Development (OECD) classification. On average, PBFS are found to increase the number of publications, though the effect is only temporary and fades after a few years. Looking at the scientific impact, PBFS are found to have a negligible effect on excellence as measured by the share of articles published in the top journals, irrespective of the type of assessment adopted. On the contrary, PBFS have some influence on average research quality, as measured by the number of citations per paper normalised with respect to the field.

## 1 | INTRODUCTION

Since the 1980s, a widespread wave of reforms has led to the redesign of the main public administration mechanisms at all levels and in all sectors (Kettl, 2005). According to Pollitt and Bouckaert (2011), one of the main drivers has been a decreasing public trust in government actions, which stimulated a growing emphasis on management, a widespread use of performance indicators and growing integration among different levels of administration and society (e.g., in the healthcare system, integration among hospitals, other healthcare centres and private research). The research and Higher Education sectors have also been heavily involved in this wave of reform. In post-Second World War Europe, public universities were mostly financed through a single block grant (a sort

of General University Fund<sup>1</sup>), provided by central governments. Research institutions also received competitive funding from the private sector or from research councils and other funding organisations. Starting from the 1980s, as a reaction to the changing perception and the strained public budget, the share of competitive funding increased (Lepori et al., 2007); the mechanism used to distribute institutional block funding started to change as well, from being mainly based on historical data to the adoption of performance indicators. In this context, performance-related financing systems (henceforth performance-based funding systems (PBFS)) started to be introduced, in order to strengthen the accountability of research institutions and steer their strategic behaviour, provide strategic information for research policies and enhance the overall quality of national research systems (Mahieu & Arnold, 2015). While not immune from some risks (see, i.e., Hicks, 2012), PBFS are indeed considered to stimulate research organisations to improve their production with regards to the quality or quantity, as well as to steer certain fields or topics of research (Geuna & Martin, 2003). Further expected benefits include providing incentives and channelling resources to best-performing institutions, in a competitive game which rewards outputs (Herbst, 2007), while also improving the public perception via increased accountability (Hicks, 2012). While performance management reforms reshape incentives to be more results oriented, they increase autonomy and discretion in order to bring decisions at the most appropriate level of responsibility.<sup>2</sup>

Possible drawbacks in the adoption of such systems include the risk of creating perverse incentives for bad research practices (ranging from the multiplication of irrelevant publications to plagiarism, self-plagiarism and scientific fraud (Hazelkorn, 2010)), as well as the risk of strengthening the so-called Matthew effect and that of discouraging interdisciplinary and innovative research (Hicks, 2012; Rafols, Leydersdorff, O'Hare, Nightingale, & Stirling, 2012; Wilsdon et al., 2015). Costs of implementation of PBFS are also non-negligible (De Boer et al., 2015; Martin, 2011; OECD, 2010), though Geuna and Piolatto (2016) indicate that at least in the short term the benefits outweigh the costs in the case of the United Kingdom (UK) and Italy (see also Jonkers & Zacharewicz, 2016).

In this paper, we aim to contribute to this literature in a cross-country perspective, by estimating the impact of PBFS on the quantity and quality of research in a panel of 31 countries. According to our reading of the existing literature, it has not yet reached a consensus on the results obtained from the introduction of such systems. Research output (often measured by the number of papers) has generally been found to increase from a quantitative point of view (Moed, 2005), but it is difficult to separate the effect of PBFS from other potential drivers. As for the effect on research quality and impact, the evidence is rather mixed and there is no robust consensus (see, i.e., Adams & Gurney, 2010 or Wang & Hicks, 2013 among others). In this respect, we improve on the existing literature in three ways: (a) we expand the number of countries under analysis, some of which have not adopted a PBFS, allowing us to use them as control cases in a difference-in-difference analysis; (b) we try to characterise a different design of PBFS, arguing that the use of pure peer review systems may have a different impact when compared to systems that make use of bibliometric indicators (see also European Commission, 2017); and (c) we investigate whether or not the effects of PBFS are homogenous across different research fields. The rest of the paper is structured as follows: section two introduces a simple model intended to highlight potential reasons for the effectiveness of performance-related incentives, while section three provides a definition of PBFS in the context of financing of the university system and presents the dataset; section four discusses our empirical strategy and section five presents our results. Finally, section six concludes.

## 2 | A THEORETICAL FRAMEWORK

### 2.1 | A simple model

In a standard agency model (Lazear & Gibbs, 2014), whenever output is imperfectly observable from a principal and effort is costly to an agent, there is room for incentive-compatible contracts. Under risk neutrality of both contracting partners, the typical solution suggests that the principal should reward the agent according to the

observed performance, leaving the enterprise risk on the shoulders of the agent. On the contrary, when the agent is risk averse, the principal should reduce the indexation of the reward to performance. In this literature, fixed rewards are obtained as limiting cases of extreme volatility to the observable output (the signal-to-noise ratio goes to zero) and/or extreme risk aversion of the agents (when the principal provides an implicit insurance contract to the agent). In addition, in the case of multiple outputs, the reward should be made proportional to all outputs (in such a way that the marginal benefits net of costs be equalised across tasks), in order to avoid the agent investing all energies in the most rewarding activity. Can we take these suggestions to the case of a university manager confronting an academic professor? More generally, are these models applicable to the framework of Higher Education institutions?

Let us suppose to optimally design an incentive-compatible contract for an academic who is performing two types of activities, teaching (measured by  $x$  hours spent with students) and research (which could be proxied by  $y = \sum_{i=1}^n q_i = n \cdot \bar{q}$ , where  $n$  is the number of published papers weighed by  $q_i$ , the relevance of the journal where each paper is published). Then the optimal wage contract would be

$$w = \alpha + \beta x + \gamma n \bar{q}. \quad (1)$$

If effort  $e$  is required in performing each type of activity, an academic will choose the optimal amount of effort by maximising her utility function given by

$$U = w - \delta \log(x) - \eta \log(n \bar{q}) \quad (2)$$

where the second and third addends capture the disutility of effort required to perform each activity.<sup>3</sup> Notice that we are implicitly assuming that teaching and research are perfect substitutes in the subjective perception of the agent. Maximising Equation 1 subject to Equation 2 leads to

$$x^* = \frac{\beta}{\delta}; (n \bar{q})^* = \frac{\gamma}{\eta}. \quad (3)$$

In this specific formulation, the research activity is positively associated with the specific reward parameter  $\gamma$ , without negative spill over effects onto teaching activity. If there is some heterogeneity across academics with regards to subjective cost of research (parameter  $\eta$ ), then the sensitivity to incentives will be higher among those with lower costs.

This set-up can be used to frame the potential consequences of introducing/strengthening a performance-related system of rewards. According to Equation 3, other things being constant, one should observe an increase of effort devoted to research. However, this simple model does not tell us whether or not the increase of research activity will imply an increase in the number of papers for a given impact (with regards to citations, impact factor, etc.) or an improvement in the average impact (for a given paper productivity).

## 2.2 | Plausibility of theoretical predictions in a micro perspective

The predictions of the model proposed in the previous section are subject to several caveats that we will discuss in the present paragraph. First of all, the research activity is subject to high uncertainty with regards to final outcome (will one get the expected result?, will one be able to publish in highly ranked journals? and so on). As a consequence, despite common practice in top departments, it would not be advisable to tie pay to research performance. In addition, most academics obtain wage increases by either being promoted or by moving to other departments, a dimension which is neglected in this simple model.

Second, academics respond to non-monetary incentives as well, like prestige, visibility, fame among peers, curiosity and the like. Even in the absence of monetary rewards, one could expect research not being nil due to

these other incentives. When more weight is shifted to monetary rewards it is not clear what may occur to non-monetary incentives: they could either decline/disappear (thus rendering the overall effect on research negligible or even negative) or become stronger (in a sort of deontological reaction against the monetary evaluation of their internal mission).

Third, research activity is also a team production, in some research fields more than in others. Individual performance-related pay exacerbates competition and reduces cooperation (especially if the reward takes the form of a tournament). Thus, unless the reward policy is converted into group rewards associated with group productivity, there is a risk associated with these incentive policies since they can yield a reduction in the overall productivity.

These risks are openly discussed, among others, by Moynihan and Pandey (2010) and Rabovsky (2014). The former paper argues that performance indicators are not actually used in public administration management for incentive purposes.<sup>4</sup> Whenever they are used, it is rather for organisational purposes (extrinsic motivation of public employees, their organisational commitment, their job satisfaction and involvement). Managers that are more likely to use performance information are those pursuing decision flexibility and/or public accountability of public services. All in all, the authors do not attribute a significant incentive role to performance-related indicators. The second paper analyses a survey among US public college administrators, showing that incentive reasons are not among the top ones in adopting performance funding. Rather ideology (partisanship of the state legislature, political orientation of the administrator) or organisational learning/improvement seem to drive the utilisation of these management practices.<sup>5</sup> The assessment of the effectiveness of these practices is rather discouraging, possibly due to the multiplicity of goals that Higher Education institutions are contemporaneously pursuing.

### 2.3 | Different levels of aggregation for validating this theory

Previous implications are referred to an individual contract between a principal (e.g., a university dean) and an agent (e.g., a university professor). Wherever academic salaries are individualised (as in the Anglo-Saxon world), employment contracts associate part of the payment to research productivity (De Fraja, Facchini, & Gathergood, 2016). In such a framework, the impact of a funding reform could be tested at the individual level, by comparing the scientific production of the same individual before and after the change in the rule. However, in countries where salaries are homogeneous within academic ranks, an individual does not receive a personal benefit from putting greater effort into research activity since research performance is typically measured at an aggregate level (department, university or region) and the prizes are paid at the same level of aggregation.<sup>6</sup>

If we reformulate the previously sketched model with regards to group incentives, we can add a further dimension that we have not considered so far: the free-riding incentive. Whenever an output is jointly measured with reference to a specific group of agents, the members of the group face the typical dilemma between fully contributing to the output while obtaining only a fraction of the returns. The standard prediction of a team incentive model is that no agent will put effort into producing the output, unless punishment strategies (by the principal or by peers) are considered. Thinking of university departments, the blame and shame strategy is perhaps the most stringent incentive to conduct research activity, even in the absence of personal rewards. This leads us to consider the potential impact of performance rewarding on an aggregate entity (be it a department or even a university). As long as internal lines of responsibilities are clear enough and accepted, a performance assessment may elicit more effort from the group, especially if it is framed with regards to a tournament among groups. The performance assessment creates a competition among participating entities, and the internal faculties will strive to achieve a larger share of the aggregate funding. In such a case, we expect reforms of the funding systems to be quite effective in raising research activity, unless the excess of competition destroys the cooperative attitude prevailing in many research fields. One could empirically test these predictions at department or university level, but should pay attention to many confounding factors that hamper the identification of a pure incentive effect on research activity.<sup>7</sup>

These confounding factors tend to cancel each other if we go up the level of aggregation, by considering the impact of performance assessment on the research activity of the national academic community. Since each

individual academic would be unable to compute personal costs and benefits, and the competition among departments/universities would be attenuated the larger the number of competitors, when considering the country level one would be investigating the effectiveness of governments to govern and steer Higher Education systems as they do for other publicly funded services (Ferlie, Musselin, & Andresani, 2008). Here the problem is given by the multiplicity of instruments that a government has available, especially those connected to hiring, motivating and remunerating the research bodies. Seen in this final perspective, the research design cannot help being cross-country comparative. Its main challenge is how to get rid of many other policy changes that may have taken place in the same period that may also have affected the ability to conduct research activities. In the sequel we take these problems into account by netting out country and year fixed effects, as well as a common time trend that may underlie the variables' dynamics. The remaining variation in the data reflects the policy changes we are interested in. The underlying assumption is that research activity is fully internationalised and therefore subject to common shocks across countries. As a consequence, if one observes an increase in research in country A (and not in country B), while with similar timing one also records the introduction of PBFS in country A (and not in country B), one can safely claim that the policy reform has impacted on research activity. In such a case, country B acts as counterfactual for country A since it represents the research activity that would have been observed if PBFS had not been introduced.

### 3 | PERFORMANCE-BASED FUNDING SYSTEMS: DEFINITION AND DATASET

#### 3.1 | Defining performance-related funding systems

According to the definition formulated by Hicks (2012), and operationalised in Jonkers and Zacharewicz (2016), the main characteristics of a PBFS are as follows:

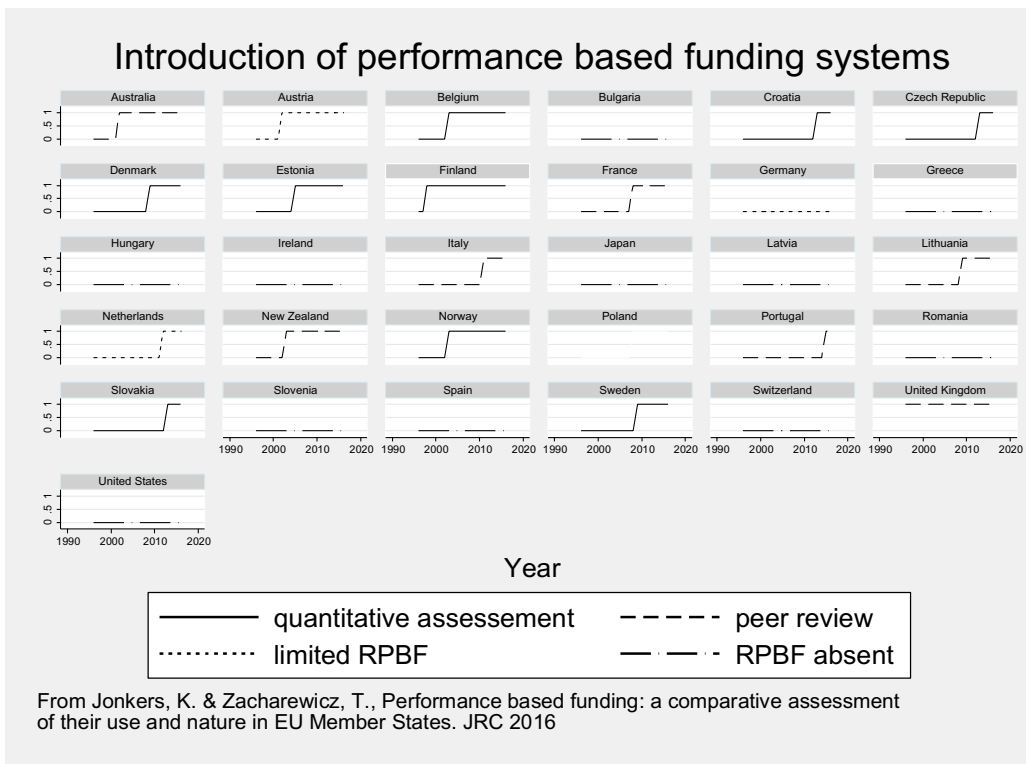
1. Research must be assessed.
2. Research evaluation must be ex post.
3. Research output and/or impact must be evaluated.
4. Part of the governmental allocation of university research funding must depend on the outcome of the evaluation.
5. The PBFS must be a national or regional system.

Assessment exercises concerning degree programmes and teaching are excluded. Also, evaluations of projects' proposals are not included, because they are mainly performed ex ante, and not directly involved with the outcomes of research activity derived from those proposals. For the same rationale, funding systems that reallocate funds only on the number of PhD students are not considered. Moreover, a PBFS must contribute directly or indirectly to assign research funds: any kinds of evaluation exercises that provide only recommendations or feedback to Higher Education institutions and Public Research Organisations (PROs) are not considered, as well as evaluation or assessment exercises that are performed at local or institutional level (Hicks, 2012).

Based on this definition, Jonkers and Zacharewicz (2016) provide a general overview of the different assessment and allocation systems in use among the European Union Member States and in some of the main extra-European countries. Among the latter, in the United States the Higher Education system is mainly funded at the institutional level and is not usually considered as a PBFS according to the Hicks definition, while Australia and New Zealand have experienced several evaluation exercises in the last 20 years and hence are considered as proper PBFS systems (Butler, 2003a, 2003b). This classification takes into account two kinds of information: the presence of some types of PBFS, according to Hicks' definition reported above, and the year in which this kind of

system has been introduced in the country. They also distinguish between quantitative assessment systems, often based on bibliometric indicators, and informed peer review systems. The former is often automatic, operates on a regular basis (in some cases even with annual frequency), while the latter is more irregular with regards to timing, due to the higher costs of implementation.<sup>8</sup> The former system is expected to exert more pressure on researchers in publishing since this is the more easily monitored output dimension. The effects on quality of research cannot be attributed to one system more than another, for quantitative assessments are often combined with analysis of impact measures, whereas peer review systems are built on (and simultaneously build) reputation, which in turn yields visibility.<sup>9</sup> Their classification was enriched by the identification of the presumed date of impact of the different PBFS, though we are fully aware of the limits that setting a specific date of impact implies (Hicks, 2017). In fact, it is not always clear which policies may be considered as the first introduction since in some countries multiple policy changes occurred over the span of time we examined (as in the case of Australia).<sup>10</sup> Secondly, it is not simple to single out a precise date for policy implementation, especially when searching for impact on research activities (Martin & Whitley, 2010; Wang & Hicks, 2013). Changes in funding strategies are often concomitant to other policy changes, which may also affect research activities (i.e., recruitment policies—Hicks, 2017). We cope with these problems by distinguishing between different types of evaluation systems according to the most detailed and updated overview available (Jonkers & Zacharewicz, 2016). As regards the date of implementation, we have chosen the date of introduction of new rules for the performance-based funding system or (when available) the foundation dates of national agencies responsible for the evaluation of research activities, whose indicators are then used to measure performances.

Figure 1 and Table 1 show the classification of the 31 countries considered in our analysis<sup>11</sup> for the period 1996–2016, distinguishing between the two types of systems (and accounting for intermediate cases). The figure



**FIGURE 1** Presence and type of PBFS across countries and years

**TABLE 1** Classification of country experiences

| Country   | Type of PBFS                                 | Years of introduction |
|---|--|-----------------------|
| <i>Absence of research performance-based funding</i>                              |  |                       |
| Bulgaria  | No PBFS                                      | na                    |
| Greece  | No PBFS                                      | na                    |
| Hungary   | No PBFS                                      | na                    |
| Ireland   | No PBFS                                      | na                    |
| Japan   | No PBFS                                      | na                    |
| Latvia  | No PBFS                                      | na                    |
| Romania   | No PBFS                                      | na                    |
| Slovenia  | No PBFS                                      | na                    |
| Spain   | No PBFS                                      | na                    |
| Switzerland   | No PBFS                                      | na                    |
| United States   | No PBFS                                      | na                    |
| <i>Funding not based on research performance but on other teaching indicators</i> |  |                       |
| Austria   | Limited PBFS                                 | 2002                  |
| Germany   | Limited PBFS                                 | na                    |
| Netherlands   | Limited PBFS                                 | 2012                  |
| <i>Funding based on research performance measure through peer review</i>          |  |                       |
| Australia   | Peer review                                  | 2002                  |
| France  | Peer review                                  | 2008                  |
| Italy   | Peer review                                  | 2011                  |
| Lithuania   | Peer review                                  | 2009                  |
| New Zealand   | Peer review                                  | 2003                  |
| Portugal  | Peer review                                  | 2015                  |
| UK  | Peer review                                  | Before 1996           |
| <i>Funding based on research performance measure through bibliometric methods</i> |  |                       |
| Belgium   | Quantitative formula—bibliometric assessment | 2003                  |
| Croatia   | Quantitative formula—bibliometric assessment | 2013                  |
| Czech Republic  | Quantitative formula—bibliometric assessment | 2013                  |
| Denmark   | Quantitative formula—bibliometric assessment | 2009                  |
| Estonia   | Quantitative formula—bibliometric assessment | 2005                  |
| Finland   | Quantitative formula—bibliometric assessment | 1998                  |
| Norway  | Quantitative formula—bibliometric assessment | 2003                  |
| Poland  | Quantitative formula—bibliometric assessment | 2008                  |
| Slovakia  | Quantitative formula—bibliometric assessment | 2013                  |
| Sweden  | Quantitative formula—bibliometric assessment | 2009                  |

Main source. Jonkers and Zacharewicz (2016)—see also note 7.

shows that according to the available information, 18 out of 31 countries have introduced some form of PBFS over time; some of the major players in the research scenario (including the United States, Japan and Germany) are not considered to use a proper form of PBFS, together with some Eastern European countries. On the other

hand, the UK is the only country in which a peer review system was in place already at the beginning of the period considered in our analysis.

### 3.2 | Indicators of scientific production and impact

Our indicators of scientific production and impact are extracted from the SciVal database and are referred to the period 1996–2016. We consider 31 countries worldwide, and distinguish among six main scientific areas, following the Field of Science and Technology (FOS) classification (see Table 2).<sup>12</sup> For each country, scientific area and year we consider an indicator of scientific production (the number of publications), one indicator of scientific impact (the field-weighted citation impact (FWCI)<sup>13</sup>) and one indicator of excellence of scientific production (the share of the paper in the top 10% of journals with regards to citations<sup>14</sup>). Figure 2 presents the trend of total publications in the period; the data show a marked increase in the number of scientific outputs indexed in Scopus, a common finding in the literature, partly attributable to the enlargement of the number of journals covered by this database.

Figure 3, on the other hand, offers a picture of the impact of research with regards to citations by country and year. The FWCI indicator allows the number of citations received from the publications of a given research unit (in our case, the reference units are the countries) compared with the average number of citations received from all similar publications that are part of the bibliometric data universe. In the same period (1996–2016), European countries increased their scientific impact as measured by the FWCI, while the entire OECD remained almost constant over the sample period (see top-left panel of Figure 3). As an example, consider the case of Italy and Spain: at the start of the sample period they were lagging behind with regards to scientific impact with respect to the other main industrial countries; for all, at the end of the sample period they recovered this gap and even exceeded the value of this indicator computed on the aggregate of publications of all OECD countries (see Figure 3). The FWCI indicator provides a more interesting view in comparison with the simple number of publications since it takes into account differences in scientific production, not only concerning quantity but also between disciplines with regards to different publication and citation patterns. The World value of the FWCI indicator is set for construction equal to 1.00; a FWCI of more than 1.00 indicates that the country's publications received more citations than one would expect on the basis of the global average for the same typology of publications. For the same reason a FWCI value of less than 1.00 points to a poor performance in comparison with the global average for similar publications.<sup>15</sup>

Finally, for the same countries, Figure 4 reports the share of output comprised in the top 10 per cent of the world publications concerning number of citations. All European countries are well above the world average: in particular the share of papers comprised in the top 10 per cent worldwide in 2016 was equal to 23.4 per cent in the UK, to over 20 per cent in France and Germany while around 18 per cent also in Spain and Italy, these last values being close to the ones reached for the average of the OECD countries in the same period.

### 3.3 | Science and technology indicators

In order to reduce the risk of spurious correlation when assessing the impact of the introduction of PBFS on scientific performance, we need to control for contextual effects, including the level and structure of investments undertaken by the countries in the field of science and technology. Related indicators are extracted from the OECD Main Science and Technology Indicators (MSTI) database.<sup>16</sup> We choose to concentrate our attention on the indicators concerning the number of total researchers (with regards to full-time equivalent, which are either working in the Higher Education sector or in the entire economy<sup>17</sup>) and the share of total gross research and development (R&D) expenditures with respect to gross domestic product (GDP). Figure 5 plots the (log of) two alternative proxies for the stock of researchers (information is missing for non-OECD countries): researchers working in Higher Education and researchers in the entire economy. While the two variables are highly correlated (correlation coefficient .95), they do not display strong correlation with the research outputs. In addition, they are

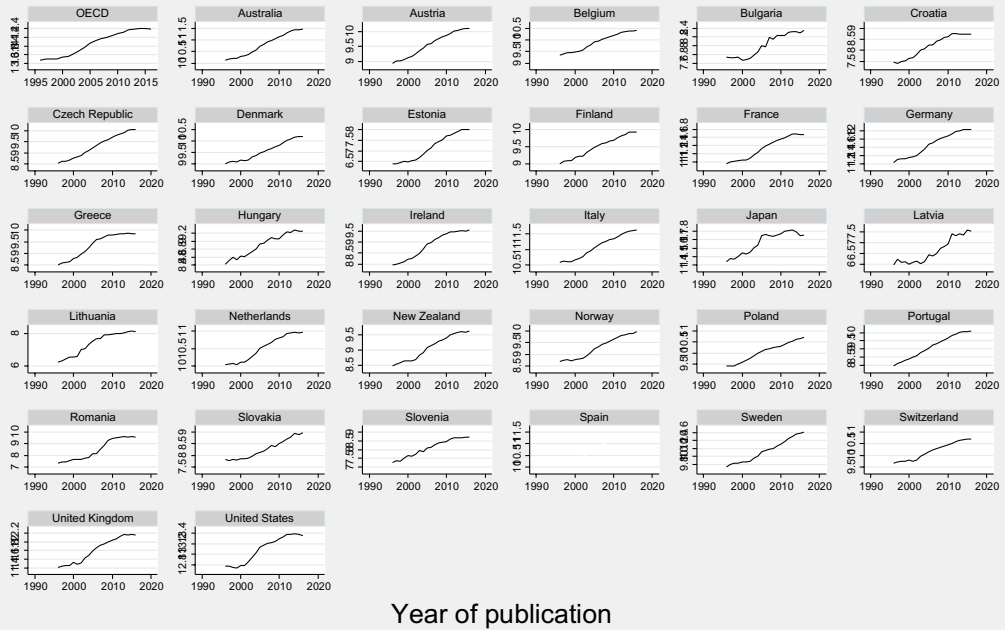


**TABLE 2** List of scientific areas

| Scientific category          | Subcategories included  |
|------------------------------|---|
| Agricultural sciences        | Agriculture, forestry and fisheries<br>Animal and dairy science<br>Veterinary science<br>Agricultural biotechnology<br>Other agricultural sciences<br>Agriculture, forestry and fisheries<br>Animal and dairy science<br>Veterinary science   |
| Engineering and technologies | Civil engineering<br>Electrical engineering, electronic engineering, information engineering<br>Mechanical engineering<br>Chemical engineering<br>Materials engineering<br>Medical engineering<br>Environmental engineering<br>Environmental biotechnology<br>Industrial biotechnology<br>Nano-technology<br>Other engineering and technologies |
| Humanities                   | History and archaeology<br>Languages and literature<br>Philosophy, ethics and religion<br>Art (arts, history of arts, performing arts, music)<br>Other humanities   |
| Medical sciences             | Basic medical research<br>Clinical medicine<br>Health sciences<br>Health biotechnology<br>Other medical science   |
| Natural sciences             | Mathematics<br>Computer and information sciences<br>Physical sciences and astronomy<br>Chemical sciences<br>Earth and related environmental sciences<br>Biological sciences<br>Other natural sciences   |
| Social sciences              | Psychology<br>Economics and business<br>Educational sciences<br>Sociology<br>Law<br>Political science<br>Social and economic geography<br>Media and communication<br>Other social sciences  |

Source. OECD (2015).

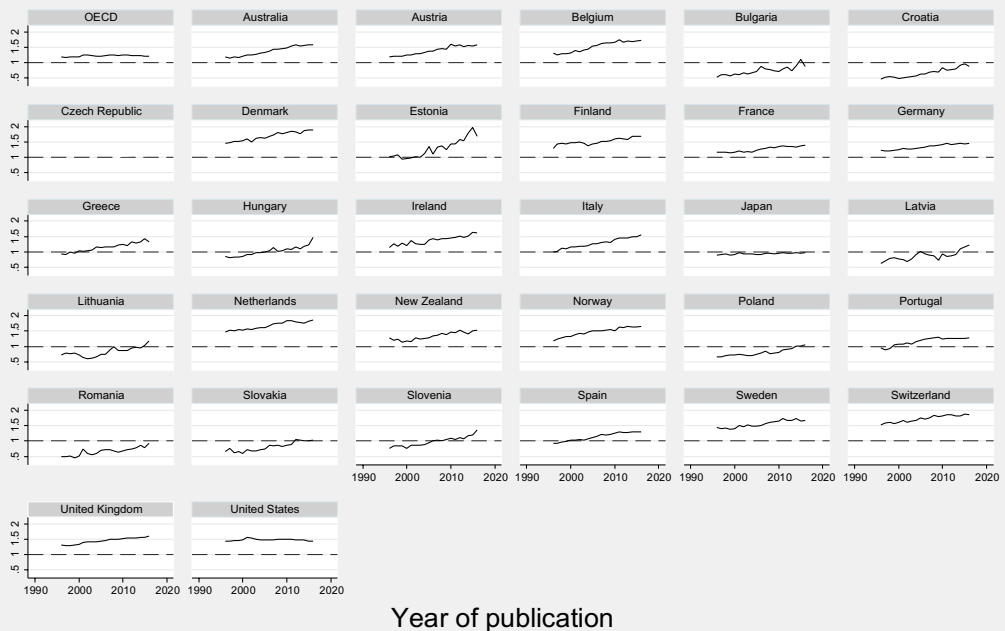
## Total number of (log)publications - rescaled by country



**FIGURE 2** Total number of publications, 1996–2016.

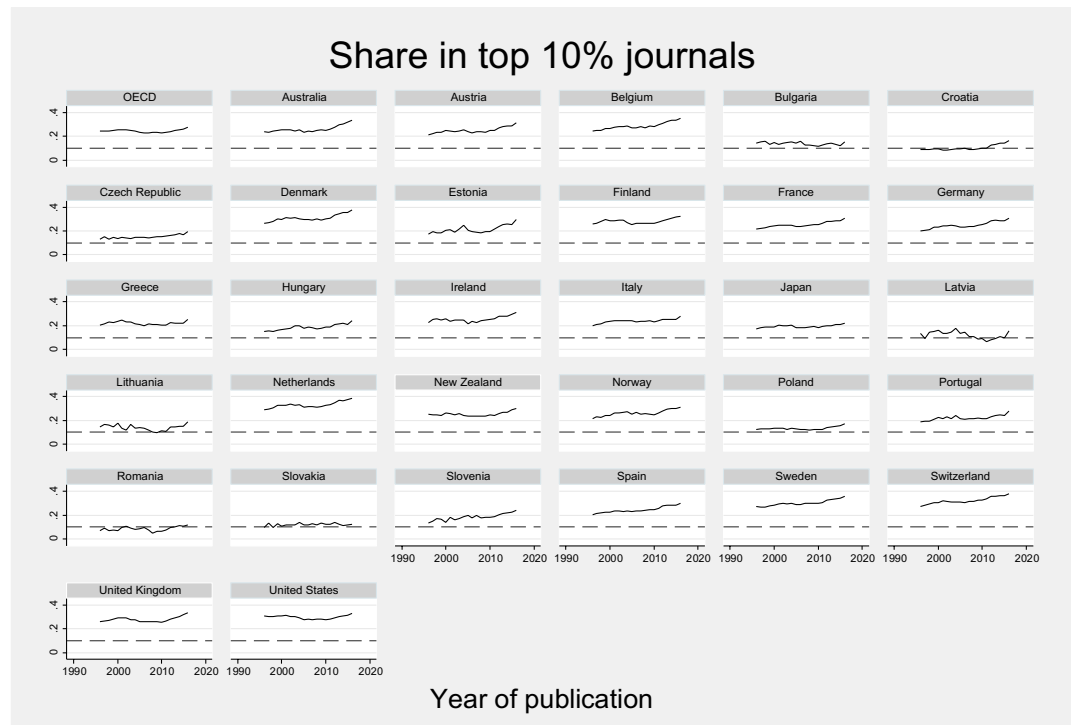
Source. Scival

## Citations over world field average



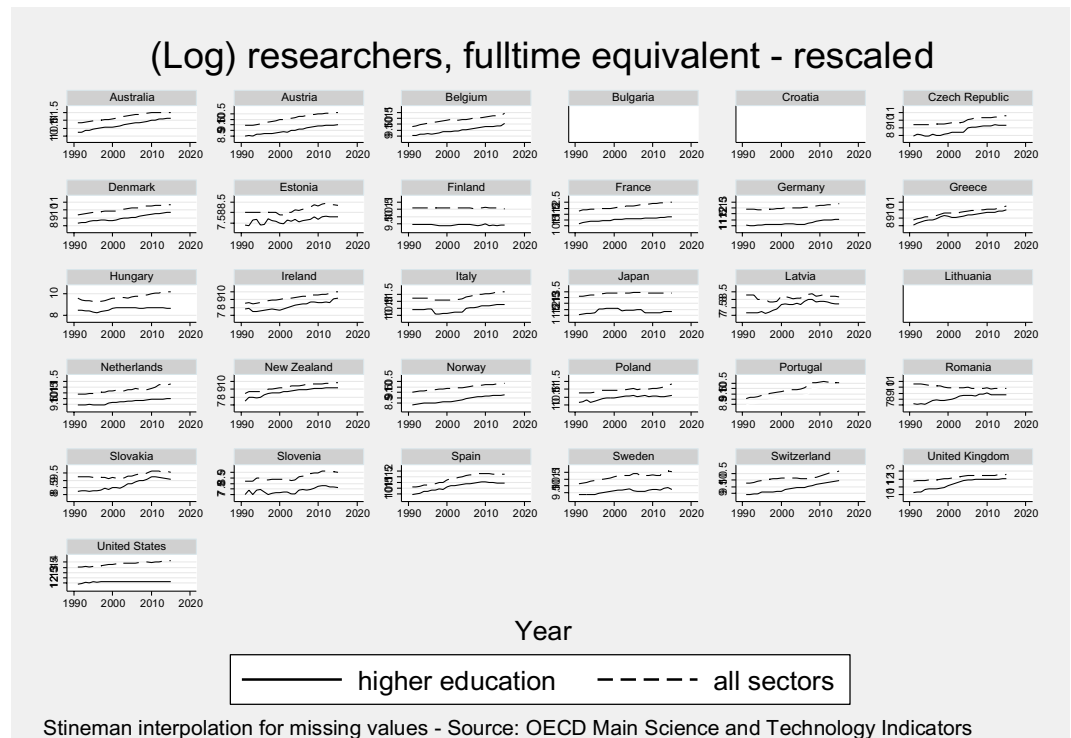
**FIGURE 3** Field-weighted citation impact, 1996–2016.

Source. Scival

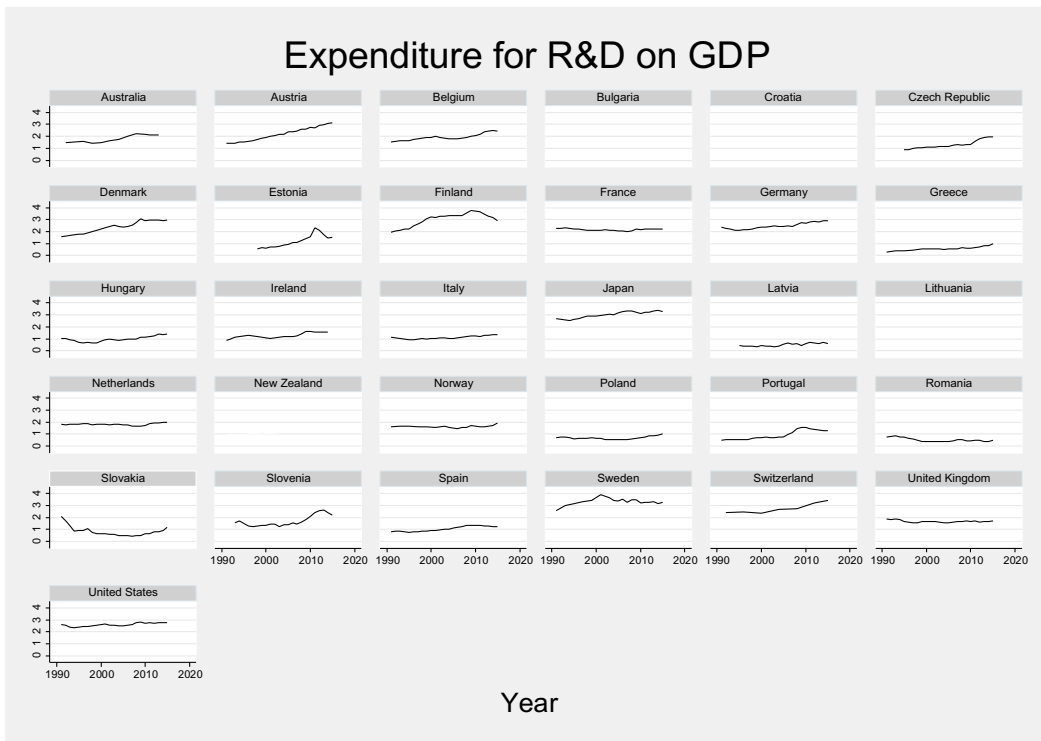


**FIGURE 4** Share of papers in the top 10 per cent in the world with regards to number of citations, 1996–2016.

Source: Scival



**FIGURE 5** Total researchers, 1991–2015



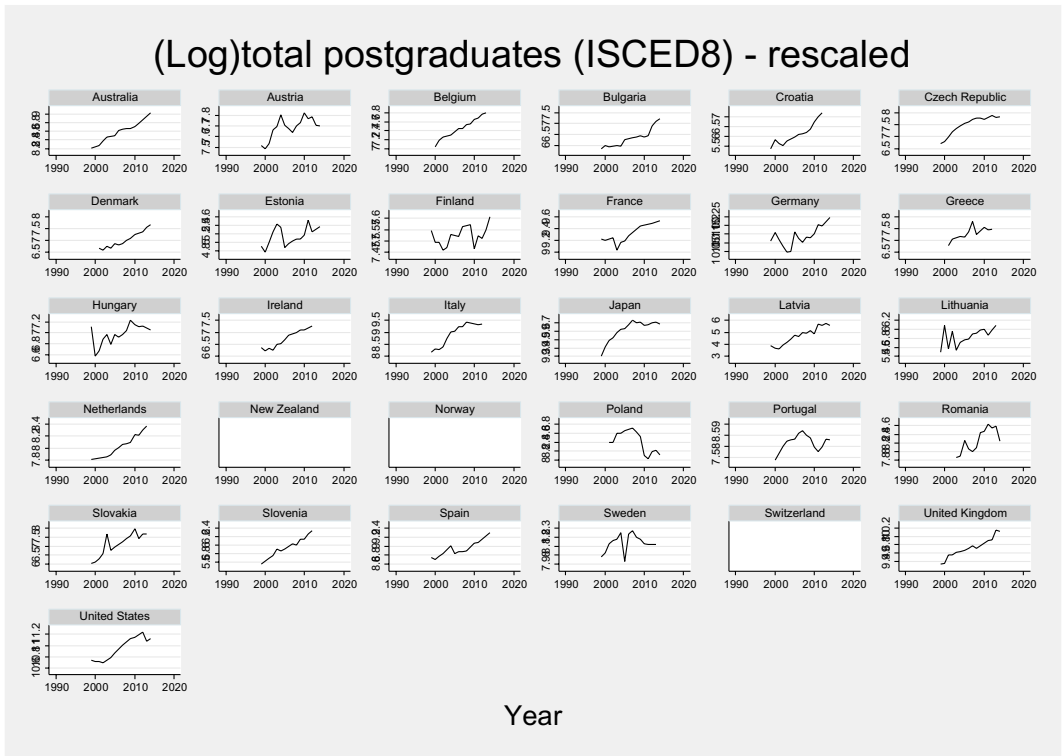
**FIGURE 6** Gross domestic expenditure on R&D (GERD) in percentage of GDP, 1991–2015.

Source. OECD

missing in almost one third of the observations, and we had to resort to data interpolation in order not to lose the excess number of degrees of freedom. In the sample period considered, France, Germany and the UK show a similar upward trend, especially when expressed as a share of the total labour force. Figure 6 shows the temporal dynamics of the share of gross domestic expenditures on R&D with respect to total GDP: in this case, Germany stands out, with a higher share that continues to grow even in the most recent years, reaching almost three per cent with regards to the country's GDP, while in Italy and Spain the share is just slightly over one per cent. Finally, we searched for some control on the supply side of potential researchers, and we resorted to the United Nations Educational, Scientific and Cultural Organization (UNESCO) database (concerning the number of postgraduate students (ISCED8 according to the UNESCO definition, <https://data.uis.unesco.org/>)), which, however, covers a shorter period of years (compared to the other series) and exhibits unusual fluctuations (see Figure 7). We will make use of this variable simply as additional controls, without placing much emphasis on it.

#### 4 | EMPIRICAL STRATEGY

We approach our question of interest (whether or not the introduction of PBFS had an impact on the research activity of a country) following a difference-in-difference approach since our sample includes 13 countries without any performance assessment systems, which will represent our control group. The remaining group of countries have introduced various forms of research assessment at different points in time, either based on algorithms or on peer review. These countries receive their treatment at a different point of the sample period, possibly with different intensity (since formula-based systems tend to bite more than peer-based ones). We implement our strategy by estimating the following model:



**FIGURE 7** Log of total number of postgraduate degrees, 1999–2014.  
Source. UNESCO

$$y_{it} = \alpha_i + \beta x_{it} + \gamma \cdot \text{Reform}_{it} + \delta \cdot t + \epsilon_{it} \quad (4)$$

where  $y_{it}$  indicates the output variable of country  $i$  at time  $t$ ,  $\alpha_i$  are country fixed effects,  $x_{it}$  are additional controls (like number of researchers or PhDs, expenditure in R&D),  $t$  represents a time trend<sup>18</sup> and  $\text{Reform}_{it}$  is a step-dummy variable which takes a unitary value when a PBFS is introduced. In this case, the reform would be raising the productivity from then onwards, due to the stronger pressure put on researchers (intercept effect), while retaining a common global trend. However, this model may be inadequate if the reform is modifying the rate at which publications are obtained (e.g., increasing the number of co-authorships): in such a case the reform if creating a slope effect can be tested by interacting the reform variable with the time trend as follows:

$$y_{it} = \alpha_i + \beta \cdot x_{it} + \gamma_0 \cdot \text{Reform}_{it} + \gamma_1 \text{Reform}_{it} \cdot t + \delta \cdot t + \epsilon_{it}. \quad (5a)$$

Alternatively, one may follow a non-parametric approach for modelling the temporal dynamics replacing the time trend with year dummies as in

$$y_{it} = \alpha_i + \beta \cdot x_{it} + \gamma_0 \cdot \text{Reform}_{it} + \gamma_1 \text{Reform}_{it} \cdot t + \delta_t + \epsilon_{it}. \quad (5b)$$

The overall effect of the reform using various specifications of the model represented by Equation 5a-b will be given by the sum of the intercept (coefficient  $\gamma_0$ ) and the slope effects (coefficient  $\gamma_1$ ) multiplied by the number of years since introduction.

**TABLE 3** Total publications (logs)—all fields—1996–2016

|                    | (1)                 | (2)               | (3)                 | (4)               |
|--------------------|---------------------|-------------------|---------------------|-------------------|
|                    | com.trend           | yr.dummy          | com.trend           | yr.dummy          |
| Reform             | 0.190**<br>(0.067)  | 0.135<br>(0.103)  |                     |                   |
| Reform x time      | (0.067)<br>(0.007)  | –0.009<br>(0.010) |                     |                   |
| Peer               |                     |                   | 0.272*<br>(0.122)   | 0.210<br>(0.147)  |
| Peer x time        |                     |                   | –0.016+<br>(0.009)  | –0.013<br>(0.011) |
| Quant              |                     |                   | 0.135*<br>(0.058)   | 0.070<br>(0.127)  |
| Quant x time       |                     |                   | –0.008<br>(0.008)   | –0.004<br>(0.012) |
| Some PBFS          |                     |                   | 0.111**<br>(0.035)  | 0.101<br>(0.077)  |
| Some PBFS x time   |                     |                   | –0.013+<br>(0.007)  | –0.012<br>(0.010) |
| Time               | 0.070***<br>(0.007) |                   | 0.070***<br>(0.007) |                   |
| N                  | 651                 | 651               | 651                 | 651               |
| No. of countries   | 31                  | 31                | 31                  | 31                |
| R <sup>2</sup> adj | 0.846               | 0.859             | 0.846               | 0.859             |

Note. Standard errors in parentheses; country fixed effects included; year fixed effects included in columns 2–4; clustered errors by country; statistical significance: +.10, \*.05, \*\*.01, \*\*\*.001.

## 5 | RESULTS

### 5.1 | Impact on quantity

Turning to the results, we start with the potential impact on publication. In Table 3 we present the estimates of models 5a and 5b in columns 1 and 2 where we consider a unique definition of PBFS, while in columns 3 and 4 we distinguish between quantitative–bibliometric, peer reviews and intermediate cases.<sup>19</sup> The impact of the intercept of the introduction of performance-related funding is positive, leading to a significant increase in the number of publications: taken at face value (column 1 of Table 3), it would consist of an average increase at impact of 19 per cent which would then gradually disappear at 1.2 per cent per year after the reform.<sup>20</sup> Contrary to our expectation, this is mostly driven by the countries where the assessment was based on peer review: in such a case (column 3) the average increase is in the order of 17 per cent for PBFS systems based on peer review, eight per cent in the case of quantitative assessment exercises and three per cent in the intermediate cases.<sup>21</sup> Notice that these effects are temporary since they tend to attenuate with the passage of time, as if the researchers get used to performance-related incentives.

As a robustness check we introduced some measures of resources invested in research activity, like expenditure in R&D, the total number of researchers or the number of PhDs. However, this check reduces the sample by

one fifth since information on these variables is absent in six countries.<sup>22</sup> In addition, in order not to reduce the sample size even further, we interpolated missing data<sup>23</sup> and used initial and terminal values to extend the sample size. As a consequence, more than robustness checks the results are to be taken as investigation on the correlation between research outcomes and research inputs. In Table A1 in the Appendix we find that the total number of publications is positively correlated with the (log of) total number of researchers in Higher Education institutions, without statistically significant correlation with expenditure or PhDs.<sup>24</sup> Nevertheless, our coefficients of interest are confirmed, because the introduction of research performance funding retains its positive and significant impact, at least in the case of assessments based on peer reviews.

## 5.2 | Impact on quality of publications

If quantity seems to react, let us now consider what happens to the quality of scientific production, which is captured by two outcomes, the share of papers published in top journals (10%)<sup>25</sup> and the average citations obtained by each paper (net of differences attributable to field or year of publication). In Table 4 we show the estimates for the share of papers in top journals using the model described by Equation 4: the average effect is nil, even when controlling for availability of resources (see Table A2 in the Appendix). On the contrary, the second measure of quality seems more reactive to the introduction of a performance-based system. While the share of papers in top journals is by construction trendless, we revert to estimating the model specified in Equation 5a when moving to citations per paper. Looking at column 2 of Table 5 we observe a positive and weakly statistically significant effect of PBFS on the number of citations, in the order of five percentage points (above the corresponding world average). When controlling for availability of resources we find that the statistically significant effect is driven by systems based on peer review, as we had already found for the number of publications (see Table A3 in the Appendix). Notice that in this case the proxy for financial resources (the R&D expenditure in GDP) is found significant, as well as the number of Higher Education researchers.

**TABLE 4** Share of articles in top 10 per cent of journals—all fields—1996–2016

|                    | (1)                 | (2)              | (3)                 | (4)               |
|--------------------|---------------------|------------------|---------------------|-------------------|
|                    | com.trend           | yr.dummy         | com.trend           | yr.dummy          |
| Reform             | 0.004<br>(0.006)    | 0.003<br>(0.006) |                     |                   |
| Peer               |                     |                  | –0.004<br>(0.008)   | –0.003<br>(0.008) |
| Quant              |                     |                  | 0.007<br>(0.006)    | 0.006<br>(0.007)  |
| Some PBFS          |                     |                  | 0.015<br>(0.011)    | 0.012*<br>(0.005) |
| Time               | 0.002***<br>(0.000) |                  | 0.002***<br>(0.000) |                   |
| N                  | 651                 | 651              | 651                 | 651               |
| No. of countries   | 31                  | 31               | 31                  | 31                |
| R <sup>2</sup> adj | 0.356               | 0.599            | 0.363               | 0.604             |

*Note.* Standard errors in parentheses; country fixed effects included; year fixed effects included in columns 2–4; clustered errors by country; statistical significance: +.10, \*.05, \*\*.01, \*\*\*.001.

**TABLE 5** Citations per paper relative to field world average—all fields—1996–2016

|                    | (1)                 | (2)               | (3)                 | (4)               |
|--------------------|---------------------|-------------------|---------------------|-------------------|
|                    | com.trend           | yr.dummy          | com.trend           | yr.dummy          |
| Reform             | 0.036<br>(0.023)    | 0.047+<br>(0.027) |                     |                   |
| Reform x time      | 0.001<br>(0.002)    | 0.000<br>(0.003)  |                     |                   |
| Peer               |                     |                   | 0.023<br>(0.027)    | 0.030<br>(0.027)  |
| Peer x time        |                     |                   | 0.000<br>(0.002)    | –0.000<br>(0.003) |
| Quant              |                     |                   | 0.046<br>(0.037)    | 0.057<br>(0.046)  |
| Quant x time       |                     |                   | 0.002<br>(0.004)    | 0.001<br>(0.004)  |
| Some PBFS          |                     |                   | 0.008<br>(0.005)    | 0.032<br>(0.022)  |
| Time               | 0.018***<br>(0.002) |                   | 0.018***<br>(0.002) |                   |
| N                  | 651                 | 651               | 651                 | 651               |
| No. of countries   | 31                  | 31                | 31                  | 31                |
| R <sup>2</sup> adj | 0.770               | 0.777             | 0.773               | 0.780             |

Note. Standard errors in parentheses; country fixed effects included; year fixed effects included in columns 2–4; clustered errors by country; statistical significance: +.10, \*.05, \*\*.01, \*\*\*.001.

We can summarise the overall results obtained so far by saying that the introduction of research assessment systems with implications on funding seems to raise both quantity and quality of scientific publications. The effect of quantity is nonetheless temporary and fades in the short–medium term. On the contrary, the effect on quality when statistically significant does not decline and is mainly driven by peer review systems. These effects are robust to the inclusion/exclusion of proxies for resources available to research activities.

### 5.3 | Impact by research field

Once we disaggregate the outcome variables by the six research fields indicated in Table 2, we find that most of the previously described dynamics are driven by specific research area. In Table 6 we replicate column 3 of Table 3 while disaggregating the output by research fields: one can notice that informed peer review systems have an impact on four out of six fields, with the highest values recorded for humanities and social sciences, which are likely to be the least open to international debates in many countries.<sup>26</sup> It is also comforting that some hard science fields (notably agricultural) do react to quantitative assessment exercises;<sup>27</sup> more surprising is the fact that publications in the medical science sector seem insensitive to the introduction of performance assessment exercises. Looking at statistical significance of the interactions between the reform and the time trend, one is induced to conclude that the effect of peer assessments tends to vanish over time (interaction term negative and statistically significant in five out of six cases) while quantitative assessments look more permanent (interaction term negative but never significant). It is also interesting to note that the introduction of a performance-related



**TABLE 6** Total publications (logs)—by fields—1996–2016

|                    | (1)                 | (2)                 | (3)                 | (4)                 | (5)                 | (6)                 |
|--------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
|                    | Agricultural        | Engineering         | Humanities          | Medical             | Natural             | Social              |
| Peer               | 0.345<br>(0.238)    | 0.383**<br>(0.139)  | 0.564+<br>(0.330)   | 0.222<br>(0.187)    | 0.230+<br>(0.117)   | 0.620*<br>(0.258)   |
| Peer x time        | -0.027+<br>(0.015)  | -0.019+<br>(0.010)  | -0.036*<br>(0.016)  | -0.011<br>(0.013)   | -0.016+<br>(0.009)  | -0.039**<br>(0.014) |
| Quant              | 0.178+<br>(0.103)   | 0.223*<br>(0.089)   | 0.093<br>(0.132)    | 0.120<br>(0.107)    | 0.099+<br>(0.053)   | 0.294**<br>(0.091)  |
| Quant x time       | -0.018<br>(0.011)   | -0.005<br>(0.009)   | -0.000<br>(0.015)   | -0.013<br>(0.011)   | -0.006<br>(0.007)   | -0.019+<br>(0.011)  |
| Some PBFS          | -0.045<br>(0.049)   | 0.281***<br>(0.025) | 0.382***<br>(0.088) | 0.040<br>(0.033)    | 0.048<br>(0.034)    | 0.215**<br>(0.062)  |
| Some PBFS x time   | -0.006<br>(0.011)   | -0.023*<br>(0.010)  | -0.033+<br>(0.016)  | -0.012<br>(0.008)   | -0.007<br>(0.009)   | -0.028*<br>(0.013)  |
| Time               | 0.080***<br>(0.011) | 0.070***<br>(0.007) | 0.167***<br>(0.012) | 0.070***<br>(0.009) | 0.067***<br>(0.007) | 0.135***<br>(0.011) |
| N                  | 651                 | 651                 | 650                 | 651                 | 651                 | 651                 |
| No. of countries   | 31                  | 31                  | 31                  | 31                  | 31                  | 31                  |
| R <sup>2</sup> adj | 0.725               | 0.827               | 0.875               | 0.765               | 0.843               | 0.871               |

Note. Country fixed effects included; clustered errors by country; statistical significance: +.10, \*.05, \*\*.01, \*\*\*.001.

**TABLE 7** Share of articles in top 10 per cent of journals—by fields—1996–2016

|                    | (1)                 | (2)                | (3)                  | (4)                 | (5)                 | (6)               |
|--------------------|---------------------|--------------------|----------------------|---------------------|---------------------|-------------------|
|                    | Agricultural        | Engineering        | Humanities           | Medical             | Natural             | Social            |
| Peer               | 0.004<br>(0.034)    | -0.009<br>(0.013)  | 0.001<br>(0.018)     | -0.013<br>(0.011)   | 0.005<br>(0.005)    | -0.022<br>(0.019) |
| Quant              | 0.006<br>(0.023)    | -0.019<br>(0.014)  | 0.001<br>(0.021)     | 0.025**<br>(0.008)  | 0.006<br>(0.006)    | -0.007<br>(0.012) |
| Some PBFS          | 0.082+<br>(0.045)   | 0.013<br>(0.016)   | -0.066<br>(0.071)    | 0.012*<br>(0.005)   | 0.008+<br>(0.004)   | 0.014<br>(0.010)  |
| Time               | 0.012***<br>(0.002) | 0.002**<br>(0.001) | -0.013***<br>(0.001) | 0.004***<br>(0.001) | 0.002***<br>(0.000) | -0.001<br>(0.001) |
| N                  | 651                 | 651                | 650                  | 651                 | 651                 | 651               |
| No. of countries   | 31                  | 31                 | 31                   | 31                  | 31                  | 31                |
| R <sup>2</sup> adj | 0.645               | 0.147              | 0.403                | 0.538               | 0.427               | 0.037             |

Note. Country effects included; clustered errors by country; statistical significance: +.10, \*.05, \*\*.01, \*\*\*.001.

form of monitoring produces effects in humanities, social science and engineering, and these effects are rather precisely estimated. Also, in this case, the effect seems to vanish over a 10-year period.

**TABLE 8** Citations per paper relative to field world average—all fields—1996–2016

|                    | (1)                 | (2)                  | (3)                  | (4)                 | (5)                 | (6)                 |
|--------------------|---------------------|----------------------|----------------------|---------------------|---------------------|---------------------|
|                    | Agricultural        | Engineering          | Humanities           | Medical             | Natural             | Social              |
| Peer               | 0.055<br>(0.095)    | −0.010<br>(0.070)    | −0.031<br>(0.172)    | 0.094<br>(0.072)    | −0.014<br>(0.037)   | 0.123**<br>(0.037)  |
| Peer x time        | 0.000<br>(0.006)    | 0.000<br>(0.005)     | 0.000<br>(0.007)     | −0.007<br>(0.005)   | 0.002<br>(0.003)    | −0.012**<br>(0.004) |
| Quant              | 0.101**<br>(0.028)  | 0.093<br>(0.082)     | 0.082<br>(0.113)     | 0.103+<br>(0.060)   | 0.017<br>(0.032)    | 0.138***<br>(0.030) |
| Quant x time       | −0.007*<br>(0.003)  | −0.006<br>(0.005)    | −0.002<br>(0.009)    | 0.003<br>(0.008)    | 0.001<br>(0.004)    | −0.009**<br>(0.002) |
| Some PBFS          | 0.244***<br>(0.027) | 0.192***<br>(0.027)  | −0.077***<br>(0.011) | −0.029*<br>(0.013)  | 0.047***<br>(0.005) | 0.010<br>(0.025)    |
| Some PBFS x time   | −0.015*<br>(0.007)  | −0.015***<br>(0.002) | −0.003<br>(0.007)    | −0.001<br>(0.006)   | −0.002<br>(0.002)   | 0.006<br>(0.005)    |
| Time               | 0.017***<br>(0.003) | 0.009***<br>(0.002)  | 0.004<br>(0.004)     | 0.032***<br>(0.004) | 0.015***<br>(0.002) | 0.020***<br>(0.002) |
| N                  | 651                 | 651                  | 651                  | 651                 | 651                 | 651                 |
| No. of countries   | 31                  | 31                   | 31                   | 31                  | 31                  | 31                  |
| R <sup>2</sup> adj | 0.430               | 0.194                | −0.003               | 0.560               | 0.658               | 0.330               |

Note. Country effects included; clustered errors by country; statistical significance: +.10, \*.05, \*\*.01, \*\*\*.001.

When we consider the share of papers in top journals in Table 7, we consistently find the absence of statistical significance, with the exception of medical sciences, where quantitative assessment seems to be effective in raising the publication standards of the countries affected by research assessment.

Finally, when considering the citations per paper, we replicate the estimation of the model reported in column 3 of Table 5 by research field. In line with this result, in Table 8 we find that the strongest impact is associated with quantitative measures in three fields: agricultural sciences, medical sciences and social sciences (the impact effects are 10%, 10% and 14%, respectively). On the contrary, no statistically significant effect is found for assessment based on peer review, except in the case of social sciences. It is worth noting the situation of countries with some sort of performance-based assessment, which seems to affect various research areas in different ways, but this may simply reflect the compositional effect of the two countries involved.

## 6 | CONCLUSIONS

Performance-based funding systems have been introduced in a fairly large number of countries since the mid-1990s. Their potential impact on the research system has been widely debated in the literature, but a proper assessment of their actual impact on proxies for both quantity and quality of scientific publications is still scarce. We tried to fill this gap, using information concerning the number of publications and their scientific impact with regards to citations or publications in top-ranked journals for 31 countries worldwide over the period 1996–2016. Eighteen countries have introduced some form of performance-based funding system within the period considered, while the remaining countries, that either do not distribute funding according to a measure of performance (or have started doing so before the start of our observation window), are used as control cases. The analysis has

been performed both at the aggregate level and looking separately at each of the six main scientific areas identified according to the OECD classification. A number of interesting results emerge from our analysis: on average, PBFS are found to increase the number of publications, however, the effect is only temporary and eventually dies out after a few years. The impact of PBFS on the quantity of scientific publication is particularly significant when an informed peer review system is used for the evaluation; on the other hand, evaluation based on metrics alone seems to have a weaker quantitative effect. At the level of each specific scientific discipline, the increase in the number of publications is particularly relevant in fields that are less exposed to an international audience (social sciences and humanities). Looking then at the scientific impact, we find that PBFS do not have a sizeable effect on excellence as measured by the share of articles published in the top journals, irrespective of the type of assessment adopted. On the contrary, introducing an evaluation system (especially if based on metrics) may influence average research quality, as measured by the number of citations per paper normalised with respect to the field. The effect is significant for the sectors more exposed to an international audience (medical and agricultural sciences), but it is also significant for social sciences.

We can therefore confirm one of the results already obtained from similar analysis (see, i.e., Moed, 2005), namely that starting to assess research performance and relating funding to results seems to be associated with a positive impact on research quantity, albeit on a temporary basis. Furthermore, contrary to previous findings (see, i.e., Butler, 2003b), we are also able to document a more permanent effect on research quality (as measured by citations) for a large sample of OECD countries. However, we are unable to disentangle the effect of assessing from the effect of funding, because of the lack of a sufficient number of cases where the former is adopted without the latter. Similarly, we are unable to measure the gradient of PBFS on various dimensions of research, again for lack of adequate information. Collecting data on these two dimensions is the content of our future research.

As a final consideration, one of the questions that may arise from these results is whether or not an increase in research quantity (temporary) and quality (permanent) is enough to justify the adoption of this kind of system. Indeed, the literature provides various examples of possible drawbacks from the adoption of PBFS, including providing possible disincentives for innovative and interdisciplinary research and favouring academic conformism (Rafols et al., 2012). Another possible bias induced by PBFS is that of discouraging scientific production in national languages, which has a positive effect on internationalisation of research, but may negatively affect the biodiversity of research languages, especially in humanities and the social sciences (Stöckelova, 2012). Moreover, the literature also emphasises that PBFS can be very expensive; an analysis on this aspect can be found in Geuna and Piolatto (2016), who compare PBFS costs with those incurred with financing systems based on historical costs or on ex ante competitive evaluation of research projects. Geuna and Piolatto (2016) conclude that funding systems based on performance show higher costs than those simply based on historical records, but are less expensive than those relying on ex ante evaluation. Moreover, PBFS are found to be more efficient the higher the share of funds which is distributed on the basis of the results of the evaluation.

On the basis of the results of the paper and of the possible considerations about the costs associated with PBFS documented in the literature (see also European Commission, 2017), we can conclude that PBFS are a very useful instrument to steer the university system and increase the overall impact of scientific research of a country on a permanent basis. However, particular care should be employed in the correct design of the systems, in order to minimise possible drawbacks and increase their cost efficiency.

## ENDNOTES

<sup>1</sup> 'The British higher education system until the end of the seventies provided a good example of this conception: the state allocated a public budget to the UGC (University Grant Committee), a purely academic body which then distributed it to highly collegial higher education institutions' (Ferlie et al., 2008, p. 327).

<sup>2</sup> '... performance management can be seen as an extension of the New Public Management ideology that stresses managerial creativity and adaptability as mechanisms for improving public management' (Rabovsky, 2014, p. 762).

- <sup>3</sup> Notice that we are abstracting from the issue of complementarity/substitutability between teaching and research, which would be implied by explicitly modelling the joint production of the two activities. If one is able to assume that teaching and research were perfect substitutes in the disutility cost of effort, we would end up in a corner solution (namely effort would be entirely allocated to the most remunerated activity).
- <sup>4</sup> 'Although governments have devoted a great deal of energy and resources into creating performance information systems, they have largely neglected the question of how to foster information use' (Moynihan & Pandey, 2010, p. 2).
- <sup>5</sup> 'Despite the promise and potential of these efforts, a number of recent studies have concluded that performance funding-oriented reforms have had negligible impacts on organizational performance and student outcomes' (Rabovsky, 2014, p. 763). However, it should be remembered that performance in this contribution mostly concerns teaching activities (indicators are mostly centred on graduation rates, often by population subgroups—see Table 1 therein). When confronted with research activity, teaching results are less subject to substantial uncertainty and more related to competition from similar institutions.
- <sup>6</sup> Or even higher: in the Italian case, for example, performance is measured at department level, while funds are allocated to universities according to the department performance).
- <sup>7</sup> The most notorious one is the hiring policies followed by many UK departments on the eve of the Research Excellence Framework, in order to increase their portfolios of publications to be submitted to the assessment.
- <sup>8</sup> '... countries will be classified according to the type of Performance Based Funding System they have in place, distinguishing between: (1) countries who have no performance based elements in their university funding allocation system and countries which allocate funding solely on the basis of education related metrics or assessments (without research output considerations), (2) countries which based their funding allocation formula on quantitative metrics highlighting those who use different types of bibliometric approaches and (3) countries which allocate funding on the basis of peer review based assessment exercises. The latter category can be separated into metrics based peer review and "pure peer review"' (Jonkers & Zacharewicz 2016, p. 11).
- <sup>9</sup> We would have preferred to classify funding systems by the budget share allocated to research performance and/or (even better) its gradient, but unfortunately this information could not be traced in the literature on a cross-country comparable basis. The limited evidence on the incidence of funding that is available (European Commission, 2018) indicates that this incidence ranges from 27 per cent in the UK to 1 per cent in Austria, passing through 23 per cent in Finland, 17 per cent in Italy to 7 per cent in Norway and 3 per cent in Sweden.
- <sup>10</sup> We validated our timing of reforms by consulting other sources like Butler (2003b), Auranen and Nieminen (2010), Aagaard and Schneider (2016), Geuna and Martin (2003), Franzoni, Scellato, and Stephan (2011).
- <sup>11</sup> The countries considered in the analysis are: Australia; Austria; Belgium; Bulgaria; Croatia; Czech Republic; Denmark; Estonia; Finland; France; Germany; Greece; Hungary; Ireland; Italy; Japan; Latvia; Lithuania; Netherlands; New Zealand; Norway; Poland; Portugal; Romania; Slovakia; Slovenia; Spain; Sweden; Switzerland; UK; United States.
- <sup>12</sup> See the original Frascati manual (<https://www.oecd.org/sti/inno/38235147.pdf>) and its most recent version (OECD, 2015).
- <sup>13</sup> Defined as the ratio of citations received relative to the expected world average for the subject field, publication type and publication year. We also analysed the number of citations per paper, but the variable has a clear hump-shaped dynamics over the sample period, which complicates the modelling of the dynamic path.
- <sup>14</sup> Defined as the number of publications of a selected entity that have been published in the world top 10 per cent of journals, ranked according to the CiteScore percentiles.
- <sup>15</sup> For example, a FWCI equal to 1.75 means that the country's publications have been cited 75 per cent more than what has happened at world level. On the other hand, a FWCI equal to 0.75 indicates that publications have been cited 25 per cent less than the global average. For more details on the FWCI indicator see: *Elsevier Research Intelligence SciVal Version 1.01 | February 2014 SciVal Metrics Guidebook*, available at: [https://www.elsevier.com/\\_data/assets/pdf\\_file/0020/53327/scival-metrics-guidebook-v1\\_01-february2014.pdf](https://www.elsevier.com/_data/assets/pdf_file/0020/53327/scival-metrics-guidebook-v1_01-february2014.pdf)
- <sup>16</sup> [http://stats.oecd.org/Index.aspx?DataSetCode=MSTI\\_PUB](http://stats.oecd.org/Index.aspx?DataSetCode=MSTI_PUB)
- <sup>17</sup> We also considered alternative proxies for the stock of researchers, which are available in the OECD Main Science and Technology Indicators database. In particular, we experimented with Higher Education researchers (FTE); Higher Education total R&D personnel (FTE); Government researchers (FTE); and Government total R&D personnel (FTE). The main difference between these populations concerns the precise identification and measurement of research and experimental development (R&D) personnel. OECD distinguishes between researchers who perform R&D (highly trained scientists and engineers) and other technicians and personnel (with high levels of technical experience and training) that directly support

staff in carrying out R&D projects and activities. According to the current literature, we restricted ourselves to FTE personnel, which is the main R&D statistic for international comparisons.

<sup>18</sup> The time trend accounts for the growing number of journals and publications indexed in the international databases.

<sup>19</sup> The quantitative assessment group of countries includes: Belgium, Croatia, Czech Republic, Denmark, Estonia, Finland, Norway, Poland, Slovakia and Sweden. The peer review group includes: Australia, France, Italy, Lithuania, New Zealand, Portugal and UK (see Table 1).

<sup>20</sup> Considering sample means the overall effect would be given by the result of  $(0.190 - 0.012 \cdot \overline{time}) = (0.190 - 0.013 \cdot 6.0) = 0.112$  where 6.0 is the sample mean of the time trend.

<sup>21</sup> Using the estimated coefficients in column 3 of Table 3, the effect of peer systems at sample means are obtained from  $(0.272 - 0.016 \cdot \overline{time}) = (0.272 - 0.016 \cdot 6.0) = 0.176$ , while the effect of quantitative bibliometric systems are obtained from  $(0.135 - 0.008 \cdot \overline{time}) = (0.135 - 0.008 \cdot 6.0) = 0.08$ . Finally, intermediate cases are obtained from  $(0.111 - 0.013 \cdot \overline{time}) = (0.111 - 0.013 \cdot 6.0) = 0.03$ .

<sup>22</sup> Bulgaria, Croatia and Lithuania do not possess information on expenditure and researchers, while New Zealand, Norway and Switzerland are missing the number of postgraduate degrees.

<sup>23</sup> We have followed the Stineman procedure available in the package *imputeTS* in R software: see Moritz (2018) and also Moritz and Bartz-Beielstein (2017).

<sup>24</sup> However, this result is rather sensitive to the proxies chosen for the stock of researchers: if we consider the total number of researchers in the R&D public and private sector, R&D expenditure and number of PhDs turn out statistically significant, while the number of researchers does not exhibit any correlation (available from the authors).

<sup>25</sup> We have experimented with alternative outcomes (like share of papers in top 1% or in top 5% journals), but in all cases we were unable to identify robust effects of the reform.

<sup>26</sup> When computing the overall effect at sample mean, the impact of peer review assessment on the total number of publications is 35 per cent for humanities, 38 per cent for social sciences, followed by engineering (27%) and natural sciences (13%).

<sup>27</sup> The effect at sample average is 6 per cent.

## ORCID

Daniele Checchi  <http://orcid.org/0000-0001-7466-8562>

## REFERENCES

- Aagaard, K., & Schneider, J. W. (2016). Research funding and national academic performance: Examination of a Danish success story. *Science and Public Policy*, 43(4), 518–531.
- Adams, J., & Gurney, K. (2010). *Funding selectivity, concentration and excellence—How good is the UK's research?* Oxford: Higher Education Policy Institute.
- Auranen, O., & Nieminen, M. (2010). University research funding and publication performance—An international comparison. *Research Policy*, 39(6), 822–834.
- Butler, L. (2003a). Explaining Australia's increased share of ISI publications—The effects of a funding formula based on publication counts. *Research Policy*, 32, 143–155.
- Butler, L. (2003b). Modifying publication practices in response to funding formulas. *Research Evaluation*, 12(1), 39–46.
- De Boer, H. F., Jongbloed, B., Benneworth, P., Cremonini, L., Kolster, R., Kottmann, A., ... Vossensteyn, H. (2015). *Performance-based funding and performance agreements in fourteen higher education systems. Report for the Ministry of Education, Culture and Science*. Enschede, The Netherlands: CHEPS, University of Twente.
- De Fraja, G., Facchini G., & Gathergood, J. (2016). *How much is that star in the window? Professorial salaries and research performance in UK universities* (CEPR Discussion Paper No. DP11638).
- European Commission. (2017). *MLE on performance-based funding systems. PRFS design: Policies and ambitions* (Thematic Report No. 1). Brussels, Belgium: Publications Office of the European Union.
- Ferlie, E., Musselin, C., & Andresani, G. (2008). The steering of higher education systems: A public management perspective. *Higher Education*, 56(3), 325–348.
- Franzoni, C., Scellato, G., & Stephan, P. (2011). Changing incentives to publish. *Science*, 333(6043), 702–703.

- Geuna, A., & Martin, B. R. (2003). University research evaluation and funding: An international comparison. *Minerva*, 41, 277–304.
- Geuna, A., & Piolatto, M. (2016). Research assessment in the UK and Italy: Costly and difficult, but probably worth it (at least for a while). *Research Policy*, 45(1), 260–271.
- Hazelkorn, E. (2010). Pros and cons of research assessment in *ISSC/UNESCO World Social Science Report 2010*. Paris, France: UNESCO Press.
- Herbst, M. (2007). Financing public universities. *Higher Education Dynamics*, 18, 90.
- Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, 41(2), 251–261.
- Hicks, D. (2017). What year? Difficulties in identifying the effect of policy on university output. *Journal of Infometrics*, 11(3), 933–936.
- Jonkers, K., & Zacharewicz, T. (2016). *Research performance based funding systems: A comparative assessment*. Luxembourg: Publications Office of the European Union.
- Kettl, D. F. (2005). *The global public management revolution*. Washington, DC: Brookings Institution Press.
- Lazear, E. P., & Gibbs, M. (2014). *Personnel economics in practice*. Hoboken, NJ: Wiley.
- Lepori, B., van den Besselaar, P., Dinges, M., Poti, B., Reale, E., Slipersæter, S., & van der Meulen, B. (2007). Comparing the evolution of national research policies: What patterns of change? *Science and Public Policy*, 34(6), 372–388.
- Mahieu, B., & Arnold, E. (2015). *R&D evaluation methodology and funding principles—The R&D evaluation methodology* (Final Report 1). Prague, Czech Republic: Ministry of Education, Youth & Sports.
- Martin, B. (2011). The research excellence framework and the impact agenda, are we creating a Frankenstein Monster? *Research Evaluation*, 20(3), 247–254.
- Martin, B. R., & Whitley, R. (2010). The UK research assessment exercise: A case of regulatory capture? In R. Whitley, J. Glaser, & L. Engwall (Eds.), *Reconfiguring knowledge production: Changing authority relationships in the sciences and their consequences for intellectual innovation* (pp. 51–80). Oxford, UK: Oxford University Press.
- Moed, H. F. 2005. *Citation analysis in research evaluation*. Dordrecht, The Netherlands: Springer.
- Moritz, S. (2018). imputeTS: Time series missing value imputation. R package version 2.6. Retrieved from <https://CRAN.R-project.org/package=imputeTS>
- Moritz, S., & Bartz-Beielstein, T. (2017). imputeTS: Time series missing value imputation in R. *The R Journal*, 9(1), 207–218. Retrieved from <https://journal.r-project.org/archive/2017/RJ-2017-009/index.html>
- Moynihan, D. P., & Pandey, S. K. (2010). The big question for performance management: Why do managers use performance information? *Journal of Public Administration Research and Theory*, 20(4), 849–866.
- OECD. (2010). *Performance-based funding for public research in tertiary education institutions, workshop proceedings*. Paris, France.
- OECD. (2015). *Frascati manual 2015: Guidelines for collecting and reporting data on research and experimental development*. Paris, France.
- Pollitt, C., & Bouckaert, G. (2011). *Public management reform. A comparative analysis: New public management, governance, and the neo-Weberian state*. Oxford, UK: Oxford University Press.
- Rabovsky, T. (2014). Support for performance-based funding: The role of political ideology, performance, and dysfunctional information environments. *Public Administration Review*, 74(6), 761–774.
- Rafols, I., Leydersdorff, L., O'Hare, A., Nightingale, P., & Stirling, A. (2012). How journal rankings can suppress interdisciplinary research: A comparison between innovation studies and business and management. *Research Policy*, 41, 1262–1282.
- Stöckelova, T. (2012). Immutable mobiles derailed: STS, geopolitics, and research assessment. *Science, Technology, & Human Values*, 37(2), 286–311.
- Wang, J., & Hicks, D. (2013). Detecting structural change in university research systems: A case study of British research policy. *Research Evaluation*, 22, 258–268.
- Wilsdon, J., Allen, L., Belfiore, E., Campbell, P., Curry, S., Hill, S., ... Viney, I. (2015). *The metric tide: Report of the independent review of the role of metrics in research assessment and management* (Project Report). London, UK: HEFCE.

**How to cite this article:** Checchi D, Malgarini M, Sarlo S. Do performance-based research funding systems affect research production and impact? *Higher Educ Q*. 2019;73:45–69. <https://doi.org/10.1111/hequ.12185>

APPENDIX

TABLE A1 Total publications (logs)—all fields—1996–2015

|                     | (1)                 | (2)               | (3)                 | (4)                |
|---------------------|---------------------|-------------------|---------------------|--------------------|
|                     | com.trend           | yr.dummy          | com.trend           | yr.dummy           |
| Reform              | 0.184<br>(0.122)    | 0.247+<br>(0.142) |                     |                    |
| Reform x time       | -0.013<br>(0.010)   | -0.019<br>(0.013) |                     |                    |
| Peer                |                     |                   | 0.346+<br>(0.170)   | 0.375*<br>(0.174)  |
| Peer x time         |                     |                   | -0.026+<br>(0.013)  | -0.030*<br>(0.014) |
| Quant               |                     |                   | 0.065<br>(0.069)    | 0.127<br>(0.090)   |
| Quant x time        |                     |                   | -0.002<br>(0.006)   | -0.007<br>(0.010)  |
| Some PBFS           |                     |                   | 0.038<br>(0.030)    | 0.168<br>(0.098)   |
| Some PBFS x time    |                     |                   | -0.009<br>(0.005)   | -0.018+<br>(0.010) |
| exp(R&D/GDP)        | 0.094<br>(0.116)    | 0.066<br>(0.129)  | 0.074<br>(0.117)    | 0.049<br>(0.131)   |
| log(HE researchers) | 0.346*<br>(0.146)   | 0.367*<br>(0.146) | 0.379*<br>(0.144)   | 0.395*<br>(0.147)  |
| log(PhDs)           | 0.147<br>(0.088)    | 0.118<br>(0.083)  | 0.150<br>(0.092)    | 0.122<br>(0.087)   |
| Time                | 0.050***<br>(0.008) |                   | 0.049***<br>(0.008) |                    |
| N                   | 500                 | 500               | 500                 | 500                |
| No. of countries    | 25                  | 25                | 25                  | 25                 |
| R <sup>2</sup> adj  | 0.877               | 0.885             | 0.881               | 0.888              |

Standard errors in parentheses; country fixed effects included; year fixed effects included in columns 2–4; clustered errors by country; statistical significance: +.10, \*.05, \*\*.01, \*\*\*.001.

TABLE A2 Share of articles in top 10 per cent of journals—all fields—1996–2015

|                     | (1)                 | (2)               | (3)                 | (4)               |
|---------------------|---------------------|-------------------|---------------------|-------------------|
|                     | com.trend           | yr.dummy          | com.trend           | yr.dummy          |
| Reform              | 0.003<br>(0.005)    | 0.003<br>(0.005)  |                     |                   |
| Peer                |                     |                   | 0.006<br>(0.006)    | 0.006<br>(0.007)  |
| Quant               |                     |                   | −0.000<br>(0.007)   | 0.001<br>(0.007)  |
| Some PBFS           |                     |                   | 0.009<br>(0.013)    | 0.006<br>(0.008)  |
| exp(R&D/GDP)        | 0.008<br>(0.006)    | 0.009<br>(0.006)  | 0.008<br>(0.006)    | 0.009<br>(0.006)  |
| log(HE researchers) | 0.002<br>(0.010)    | 0.008<br>(0.009)  | 0.002<br>(0.010)    | 0.008<br>(0.009)  |
| log(PhDs)           | −0.017<br>(0.011)   | −0.013<br>(0.012) | −0.017<br>(0.011)   | −0.013<br>(0.012) |
| Time                | 0.002***<br>(0.001) |                   | 0.002***<br>(0.001) |                   |
| N                   | 500                 | 500               | 500                 | 500               |
| No. of countries    | 25                  | 25                | 25                  | 25                |
| R <sup>2</sup> adj  | 0.353               | 0.556             | 0.354               | 0.556             |

Standard errors in parentheses; country fixed effects included; year fixed effects included in columns 2–4; clustered errors by country; statistical significance: +.10, \*.05, \*\*.01, \*\*\*.001.



TABLE A3 Citations per paper relative to field world average—all fields—1996–2015

|                     | (1)                 |      | (2)               |      | (3)                 |      | (4)               |      |
|---------------------|---------------------|------|-------------------|------|---------------------|------|-------------------|------|
|                     | com.trend           | B/SE | yr.dummy          | B/SE | com.trend           | B/SE | yr.dummy          | B/SE |
| Reform              | 0.036<br>(0.028)    |      | 0.043<br>(0.030)  |      |                     |      |                   |      |
| Reform x time       | 0.002<br>(0.003)    |      | 0.001<br>(0.003)  |      |                     |      |                   |      |
| Peer                |                     |      |                   |      | 0.060+<br>(0.034)   |      | 0.064+<br>(0.034) |      |
| Peer x time         |                     |      |                   |      | −0.001<br>(0.002)   |      | −0.001<br>(0.003) |      |
| Quant               |                     |      |                   |      | 0.016<br>(0.040)    |      | 0.022<br>(0.045)  |      |
| Quant x time        |                     |      |                   |      | 0.005<br>(0.004)    |      | 0.004<br>(0.004)  |      |
| Some PBFS           |                     |      |                   |      | −0.003<br>(0.010)   |      | 0.014<br>(0.023)  |      |
| Some PBFS x time    |                     |      |                   |      | 0.001<br>(0.002)    |      | 0.000<br>(0.003)  |      |
| exp(R&D/GDP)        | 0.064*<br>(0.031)   |      | 0.066+<br>(0.035) |      | 0.061*<br>(0.027)   |      | 0.063+<br>(0.031) |      |
| log(HE researchers) | 0.067<br>(0.041)    |      | 0.067<br>(0.042)  |      | 0.076+<br>(0.040)   |      | 0.076+<br>(0.041) |      |
| log(PhDs)           | 0.025<br>(0.016)    |      | 0.025<br>(0.018)  |      | 0.026<br>(0.018)    |      | 0.027<br>(0.020)  |      |
| Time                | 0.012***<br>(0.003) |      |                   |      | 0.012***<br>(0.003) |      |                   |      |
| N                   | 500                 |      | 500               |      | 500                 |      | 500               |      |
| No. of countries    | 25                  |      | 25                |      | 25                  |      | 25                |      |
| R <sup>2</sup> adj  | 0.784               |      | 0.789             |      | 0.789               |      | 0.794             |      |

Standard errors in parentheses; country fixed effects included; year fixed effects included in columns 2–4; clustered errors by country; statistical significance: +.10, \*.05, \*\*.01, \*\*\*.001.